From Image to Theory: Diagrammatic Abduction in Human–AI Mathematical Discovery

Douglas C. Youvan

doug@youvan.com www.youvan.ai

July 26, 2026

Abstract

A visual event may generate mathematics without constituting evidence for the physical interpretation it first suggests. This paper studies one such episode: a black disk drifting through a structured tuple image. The image first suggested a black hole and led to a finite algebra of matter–antimatter annihilation, radiation transport, compactness, and causal trapping. The same dark morphology then suggested the Boötes Void and led to a two-barrier model separating gravitational concentration from cosmic rarefaction. The associated physical systems are not equivalent. A black hole is characterized by concentration and causal inaccessibility; a cosmic void is characterized by underdensity and ordinary observability. Nevertheless, the visual ambiguity exposed a productive mathematical duality between inward and outward redistribution. We formalize this process as diagrammatic abduction: a diagram proposes candidate mathematical worlds, while subsequent formalization separates valid structural consequences from false ontological identification. The resulting discovery cycle is

image → morphology → analogy → formalization → correction → new mathematics.

The case illustrates a distinctive mode of human–AI collaboration in which visual perception supplies conjectural structure, artificial intelligence expands the associated formal possibilities, and mathematical verification disciplines both.

Keywords: diagrammatic abduction; mathematical discovery; human–AI collaboration; visual reasoning; analogy; metacognition; black holes; cosmic voids; tuple imagery.

Google Drive:
https://drive.google.com/drive/folders/15Kd8bUjm7zVaKdVIwbPhqjjV68aUZQvc?usp=sharing

A collaboration with OpenAI GPT-5.6 Thinking. CC4.0.

https://orcid.org/0009-0004-4206-4253

1. The Generative Event

Consider a structured mathematical image

$$I: X \longrightarrow \mathcal{P}, \qquad (1)$$

where $X$is a finite spatial domain and $\mathcal{P}$ is a rendered state space.

Suppose a bounded dark region

$$D_t \subseteq X \qquad (2)$$

moves through the image as time varies.

The immediate perceptual description is minimal:

$$\boxed{\text{a black disk moves through a structured field.}} \qquad (3)$$

This description does not uniquely determine the meaning of D_t.

The disk may represent:

$$\begin{gathered}\text{zero value,}\\ \text{masked data,}\\ \text{forbidden interaction,}\\ \text{absorbing state,}\\ \text{occluding object,}\\ \text{hidden information,}\\ \text{density deficit,}\\ \text{collapsed region.}\end{gathered} \qquad (4)$$

Thus the visual primitive is semantically underdetermined.

Let

$$\Im(D_t) \qquad (5)$$

denote the set of admissible interpretations of the observed disk.

Then generally,

$$|\ \Im(D_t)\ | > 1. \qquad (6)$$

The image supplies morphology before ontology.

Definition 1. Morphology

The morphology of a visual object is the collection of observable structural properties that do not depend on a specific physical interpretation.

For the moving disk, these include:

$$\mathcal{M}(D_t) = (\text{boundedness}, \text{darkness}, \text{motion}, \text{interior}, \text{boundary}, \text{interaction with background}). \quad (7)$$

Definition 2. Ontological interpretation

An ontological interpretation is a map

$$\Omega: \mathcal{M}(D_t) \longrightarrow \mathcal{W}, \quad (8)$$

where $\mathcal{W}$is a proposed mathematical or physical world.

Examples include

$$\Omega_{\mathrm{BH}}(D_t) = \text{black-hole-like region}, \quad (9)$$

and

$$\Omega_{\mathrm{V}}(D_t) = \text{cosmic-void-like region}. \quad (10)$$

The two interpretations act on the same morphology but produce different mathematical worlds.

Proposition 1. Visual morphology does not determine physical ontology

There exist distinct interpretations

$$\Omega_1 \neq \Omega_2 \quad (11)$$

such that

$$\Omega_1(\mathcal{M}(D_t))$$

and

$$\Omega_2(\mathcal{M}(D_t))$$

are both internally coherent.

Proof

The same bounded dark region may be interpreted as a causally hidden compact object or as an observable underdense region. These interpretations assign different physical meanings to darkness, but both are compatible with the visual facts of boundedness, contrast, and motion. Therefore morphology alone does not determine ontology. ■

2. First Abduction: The Black-Hole Interpretation

The first interpretation associated the moving dark disk with a black hole.

The initial analogical map was

$$\text{dark bounded region} \longmapsto \text{black hole.} \qquad (12)$$

Taken literally, this map is inadequate. A black hole is not defined merely by darkness or shape.

Formalization therefore replaced the visual association with structural requirements.

The resulting variables were:

$$M = \text{matter,} \qquad (13)$$
$$\bar{M} = \text{antimatter,} \qquad (14)$$
$$R = \text{radiation,} \qquad (15)$$

and

$$g = \text{causal geometry.} \qquad (16)$$

The local annihilation law was

$$M + \bar{M} \longrightarrow R. \qquad (17)$$

In finite form,

$$(m, \bar{m}, r) \longmapsto (m - a, \bar{m} - a, r + 2a), \qquad (18)$$

where

$$a = \min\{m, \bar{m}\}. \qquad (19)$$

This yielded the conservation law

$$m + \bar{m} + r = \text{constant}. \qquad (20)$$

Transport then redistributed the conserved content:

$$q_{t+1}^{-} = \mathcal{T}_g \mathcal{A}(q_t). \qquad (21)$$

A compactness functional was introduced:

$$\mathcal{C}_t(x, \rho) = \frac{2GE_t(x, \rho)}{c^4 R(x, \rho)}. \qquad (22)$$

A trapped-region surrogate was activated when

$$\mathcal{C}_t(x, \rho) \geq \mathcal{C}_{\mathrm{h}}. \qquad (23)$$

The formal architecture became

$$F_{\mathrm{BH}} = \mathcal{G} \circ \mathcal{K} \circ \mathcal{T}_g \circ \mathcal{A}. \qquad (24)$$

The image had not contained equations (18)–(24). It had suggested a semantic direction that formal reasoning subsequently replaced with explicit mathematical structure.

Definition 3. Abductive expansion

An abductive expansion is a map

$$\mathcal{A}_b\text{:visual cue} \longrightarrow \text{candidate formal system}. \qquad (25)$$

The first abductive expansion was

$$\mathcal{A}_b^{\mathrm{BH}}(D_t) = F_{\mathrm{BH}}. \qquad (26)$$

Remark 1. The productive error

The naive association

$$\text{black disk} = \text{black hole} \qquad (27)$$

was physically false as an identity.

Nevertheless, it was mathematically productive because correcting it required distinctions among:

$$\text{darkness, energy conversion, concentration, causal restriction.} \qquad (28)$$

The correction generated more structure than the initial analogy.

3. Second Abduction: The Cosmic-Void Interpretation

The same dark disk later suggested a cosmic void.

The second analogical map was

$$\text{dark bounded region} \longmapsto \text{underdense cosmic region.} \qquad (29)$$

Again, literal visual similarity was insufficient.

A cosmic void is not defined by optical darkness. It is defined by density relative to a background mean.

The corresponding density contrast was

$$\delta_t(x,\rho) = \frac{\rho_t(x,\rho)}{\bar{\rho}_t} - 1. \qquad (30)$$

A void-classified region satisfied

$$\delta_t(x,\rho) \le \delta_{\mathrm{v}}, \delta_{\mathrm{v}} < 0. \qquad (31)$$

The dynamical condition required net outward transport:

$$\Phi_{t,B} > 0. \qquad (32)$$

The regional energy balance was

$$E_{t+1}(B) - E_t(B) = -\Phi_{t,B}. \qquad (33)$$

Thus the missing interior energy was deposited elsewhere:

$$\Delta E(B) + \Delta E(X \setminus B) = 0. \qquad (34)$$

The formal architecture expanded to

$$F_{\mathrm{V}} = \mathcal{V} \circ \mathcal{K}_{\mathrm{V}} \circ \mathcal{T} \circ \mathcal{A}. \qquad (35)$$

The combined two-barrier theory became

$$F = \mathcal{B} \circ \mathcal{K} \circ \mathcal{T}_g \circ \mathcal{A}, \qquad (36)$$

with

$$\mathcal{B}(\delta, \mathcal{C}) = \begin{cases} H, & \mathcal{C} \geq \mathcal{C}_{\mathrm{h}}, \\ V, & \mathcal{C} < \mathcal{C}_{\mathrm{h}} \text{ and } \delta \leq \delta_{\mathrm{v}}, \\ I, & \text{otherwise.} \end{cases} \qquad (37)$$

Here

$$H = \text{trapped}, V = \text{void}, \qquad I = \text{intermediate}. \qquad (38)$$

The same dark morphology had now generated two mathematically incompatible physical interpretations but one coherent dual formalism.

Proposition 2. Black-hole and void interpretations are physically inequivalent

A black-hole-like region is characterized by

$$\mathcal{C} \geq \mathcal{C}_{\mathrm{h}}, \qquad (39)$$

whereas a void-like region is characterized by

$$\delta \leq \delta_{\mathrm{v}}. \qquad (40)$$

These conditions are logically independent.

Proof

Compactness depends on enclosed energy relative to radius:

$$\mathcal{C} \propto \frac{E}{R}. \qquad (41)$$

Density contrast depends on density relative to the mean:

$$\delta = \frac{\rho}{\bar{\rho}} - 1. \qquad (42)$$

Equal density contrasts may occur at unequal compactness, and low compactness does not imply underdensity. Therefore neither criterion determines the other. ■

The darkness of the two regions has opposite meanings:

black hole	:concentration and inaccessible interior,
cosmic void	:rarefaction and observable interior.

(43)

The physics is disconnected.

The mathematical suggestion is not.

4. Morphological Branching

Let

$$\mathcal{O}$$

denote the space of observations,

$$\mathcal{M}$$

the space of extracted morphologies, and

$$\mathcal{W}$$

the space of candidate mathematical worlds.

A morphology extractor is a map

$$\mu: \mathcal{O} \longrightarrow \mathcal{M}. \qquad (44)$$

For the moving dark disk,

$$\mu(I_t) = (\text{bounded}, \text{dark}, \text{moving}, \text{interior-bearing}, \text{background-disrupting}). \quad (45)$$

An interpretation map is not generally single-valued. It is more accurately written

$$\Omega: \mathcal{M} \longrightarrow \mathcal{P}(\mathcal{W}), \quad (46)$$

where $\mathcal{P}(\mathcal{W})$ is the power set of candidate worlds.

Thus

$$\Omega(\mu(I_t)) = \{W_{\mathrm{BH}}, W_{\mathrm{V}}, W_{\mathrm{mask}}, W_{\mathrm{abs}}, \dots\}. \quad (47)$$

Here

$$W_{\mathrm{BH}} = \text{compact trapped-region world}, \quad (48)$$
$$W_{\mathrm{V}} = \text{underdense void world}, \quad (49)$$
$$W_{\mathrm{mask}} = \text{missing-data world}, \quad (50)$$

and

$$W_{\mathrm{abs}} = \text{absorbing-state world}. \quad (51)$$

Definition 4. Semantic branching degree

For a morphology $m \in \mathcal{M}$, define

$$b(m) = |\ \Omega(m)\ |. \quad (52)$$

A morphology is semantically rigid when

$$b(m) = 1. \quad (53)$$

It is semantically productive when

$$b(m) > 1. \quad (54)$$

The moving black disk had

$$b(\mu(I_t)) > 1. \quad (55)$$

Its mathematical productivity arose partly from this multiplicity.

Definition 5. Formalization map

For each candidate world $W \in \Omega(m)$, define a formalization map

$$\Phi: W \longrightarrow \mathfrak{T}, \qquad (56)$$

where $\mathfrak{T}$ is a space of mathematical theories.

Thus

$$\Phi(W_{\mathrm{BH}}) = T_{\mathrm{BH}}, \qquad (57)$$

and

$$\Phi(W_{\mathrm{V}}) = T_{\mathrm{V}}. \qquad (58)$$

The theories contain explicit:

$$\text{states,operations,invariants,thresholds,observables.} \qquad (59)$$

Definition 6. Interpretation tree

The interpretation tree rooted at an observation $I$is

$$\mathcal{T}(I) = \{I \to \mu(I) \to W \to \Phi(W): W \in \Omega(\mu(I))\}. \qquad (60)$$

For the present case,

$$\mathcal{T}(I) \supset \begin{cases} I & \to \text{dark disk} \to W_{\mathrm{BH}} \to T_{\mathrm{BH}}, \\ I & \to \text{dark disk} \to W_{\mathrm{V}} \to T_{\mathrm{V}}. \end{cases} \qquad (61)$$

The branches share a perceptual root but diverge ontologically and mathematically.

Proposition 3. Shared morphology does not imply shared theory

Let $W_1, W_2 \in \Omega(m)$with

$$W_1 \neq W_2. \qquad (62)$$

Then generally,

$$\Phi(W_1) \neq \Phi(W_2). \qquad (63)$$

Proof

Different candidate worlds assign different meanings to the same morphological properties. These meanings induce different state variables, laws, and invariants. Therefore their formalizations need not coincide. ■

For the two principal branches,

$$T_{\mathrm{BH}} \neq T_{\mathrm{V}}, \qquad (64)$$

because

$$T_{\mathrm{BH}}$$

contains causal trapping, whereas

$$T_{\mathrm{V}}$$

contains observable underdensity and outward flux.

Nevertheless, both theories contain conservative redistribution.

This shared abstraction is

$$\mathcal{R}\text{:energy distribution} \longmapsto \text{new energy distribution.} \qquad (65)$$

The branch-specific signs are opposite:

$$\begin{matrix} T_{\mathrm{BH}} & \text{:inward concentration,} \\ T_{\mathrm{V}} & \text{:outward evacuation.} \end{matrix} \qquad (66)$$

Thus the common formal core is not darkness but redistribution.

5. Invariant Extraction

An analogy becomes mathematically productive when a feature survives removal of the original ontology.

Let

$$T_1$$

and

$$T_2$$

be theories generated by two interpretations of one morphology.

Define their shared structural content by

$$\mathrm{Core}(T_1, T_2) = \mathrm{Inv}(T_1) \cap \mathrm{Inv}(T_2), \qquad (67)$$

where $\mathrm{Inv}(T)$ denotes the structural invariants of T.

For the black-hole and void theories,

$$\mathrm{Core}(T_{\mathrm{BH}}, T_{\mathrm{V}})$$

contains:

$$\begin{gathered}\text{finite spatial state,}\\ \text{conserved total content,}\\ \text{transport across boundaries,}\\ \text{regional flux,}\\ \text{threshold classification,}\\ \text{emergent geometric labels.}\end{gathered} \qquad (68)$$

The branch-specific structures differ:

$$\begin{aligned} T_{\mathrm{BH}} &: \mathcal{C} \geq \mathcal{C}_{\mathrm{h}},\\ T_{\mathrm{V}} &: \delta \leq \delta_{\mathrm{v}}. \end{aligned} \qquad (69)$$

The extracted duality is

$$\boxed{\text{concentration} \longleftrightarrow \text{rarefaction.}} \qquad (70)$$

Definition 7. Abductive invariant

An abductive invariant is a structure J satisfying

$$J \in \mathrm{Inv}(T_1) \cap \mathrm{Inv}(T_2) \qquad (71)$$

for distinct theories T_1 and T_2 generated from a common visual morphology.

In the present case,

$$J = \text{conservative redistribution across a boundary.} \qquad (72)$$

For a region B,

$$E_{t+1}(B) - E_t(B) = -\Phi_{t,B}. \qquad (73)$$

The sign of $\Phi_{t,B}$ distinguishes the branches:

$$\Phi_{t,B} < 0 \Longrightarrow \text{net inward accumulation,} \qquad (74)$$

whereas

$$\Phi_{t,B} > 0 \Longrightarrow \text{net outward evacuation.} \qquad (75)$$

The same balance law therefore contains both formal possibilities.

Proposition 4. Opposite ontologies may share one conservation law

Suppose two models differ only in the sign of regional flux. Then both may satisfy

$$E_{t+1}(B) - E_t(B) = -\Phi_{t,B}. \qquad (76)$$

Proof

The conservation equation is independent of whether the flux is inward or outward. Its sign determines whether regional energy increases or decreases. Therefore the same conservation law supports opposite redistributive regimes. ■

This yields

$$\boxed{\text{one equation} + \text{opposite signs} = \text{two mathematical worlds.}} \qquad (77)$$

The visual object therefore contributed not a physical identity but a question:

$$\text{What does a bounded dark region mean mathematically?} \qquad (78)$$

The answer was not unique.

Its nonuniqueness generated the two-barrier theory.

6. Correction as a Generative Operator

Let

$$T^{(0)}$$

be the theory suggested by an initial analogy.

For the first branch,

$$T_{\mathrm{BH}}^{(0)}\text{:black disk} = \text{black hole.} \qquad (79)$$

For the second branch,

$$T_{\mathrm{V}}^{(0)}\text{:black disk} = \text{cosmic void.} \qquad (80)$$

Both identifications are physically inadequate.

Let

$$\mathcal{C}_{\mathrm{phys}}$$

denote the relevant physical constraints.

Define a correction operator

$$\mathfrak{R}\colon \mathfrak{T} \times \mathcal{C}_{\mathrm{phys}} \longrightarrow \mathfrak{T}. \qquad (81)$$

The corrected theory is

$$T^{(1)} = \mathfrak{R}\left(T^{(0)}, \mathcal{C}_{\mathrm{phys}}\right). \qquad (82)$$

For the black-hole branch, correction required:

$$\begin{array}{cc} \text{blackness} & \mapsto \text{causal inaccessibility,} \\ \text{dark object} & \mapsto \text{trapped geometry,} \\ \text{high intensity} & \mapsto \text{compactness-dependent collapse.} \end{array} \qquad (83)$$

For the void branch, correction required:

$$\begin{array}{ccc} \text{blackness} & \mapsto & \text{low tracer density,} \\ \text{empty region} & \mapsto & \text{relative underdensity,} \\ \text{absence} & \mapsto & \text{outward redistribution.} \end{array} \qquad (84)$$

Definition 8. Corrective gain

Let

$$\kappa(T)$$

measure the number of explicit distinctions, constraints, and invariants in theory T.

Define the corrective gain by

$$G_{\mathfrak{R}} = \kappa(T^{(1)}) - \kappa(T^{(0)}). \qquad (85)$$

A correction is generative when

$$G_{\mathfrak{R}} > 0. \qquad (86)$$

In the present case, correcting the naive analogy introduced:

energy conservation,regional flux,compactness,density contrast,causal accessibility,observability. (87

Therefore

$$G_{\mathfrak{R}} > 0. \qquad (88)$$

Proposition 5. Refutation may increase theoretical structure

There exist initial theories $T^{(0)}$for which rejection of the central identification produces a corrected theory $T^{(1)}$with

$$\kappa(T^{(1)}) > \kappa(T^{(0)}). \qquad (89)$$

Proof

The initial identification may collapse several distinct concepts into one. Refuting it requires those concepts to be separated and related explicitly. The corrected theory therefore contains more distinctions and constraints than the original. ■

The productive sequence is consequently

$$\boxed{\text{analogy} \rightarrow \text{error} \rightarrow \text{correction} \rightarrow \text{structure.}} \qquad (90)$$

This differs from a discovery model in which error is merely discarded.

Here the error determines what must be distinguished.

Definition 9. Generative mismatch

A generative mismatch is a discrepancy

$$\Delta = \text{predicted identity} - \text{validated relation} \qquad (91)$$

whose resolution produces new mathematical structure.

For the present episode,

$$\Delta_{\mathrm{BH}} = \text{visible black disk} - \text{causal horizon}, \qquad (92)$$

and

$$\Delta_{\mathrm{V}} = \text{empty-looking disk} - \text{nonempty underdensity.} \qquad (93)$$

Resolving these mismatches produced the distinction

$$\boxed{\text{hidden} \neq \text{sparse.}} \qquad (94)$$

That distinction became the conceptual center of the two-barrier model.

7. Human–AI Division of Cognitive Labor

The episode involved two distinct but coupled cognitive systems:

$$H = \text{human investigator}, \qquad (95)$$

and

$$A = \text{artificial intelligence.} \qquad (96)$$

Neither system alone should be treated as the complete discoverer.

The human supplied:

$$\begin{gathered}\text{the original visual artifact,}\\ \text{recognition of aesthetic salience,}\\ \text{the black-disk observation,}\\ \text{the black-hole association,}\\ \text{the later Boötes-Void association,}\\ \text{judgment that the physics was disconnected.}\end{gathered} \qquad (97)$$

The artificial intelligence supplied:

$$\begin{gathered}\text{rapid expansion of candidate analogies,}\\ \text{formal state variables,}\\ \text{conservation equations,}\\ \text{threshold classifiers,}\\ \text{comparison with established physics,}\\ \text{explicit correction of false identities.}\end{gathered} \qquad (98)$$

Definition 10. Human proposal operator

Let

$$\mathcal{P}_H : \mathcal{O} \times \mathcal{K}_H \longrightarrow \mathcal{Q} \qquad (99)$$

map an observation and human background knowledge $\mathcal{K}_H$ into a conjectural prompt or question $q \in \mathcal{Q}$.

For the first branch,

$$\mathcal{P}_H(I, \mathcal{K}_H) = \text{“Does this resemble a black hole?”} \qquad (100)$$

For the second branch,

$$\mathcal{P}_H(I, \mathcal{K}_H) = \text{“How does this relate to the Boötes Void?”} \qquad (101)$$

Definition 11. AI expansion operator

Let

$$\mathcal{E}_A : \mathcal{Q} \times \mathcal{K}_A \longrightarrow \mathcal{P}(\mathfrak{T}) \qquad (102)$$

map a question and machine-accessible knowledge $\mathcal{K}_A$into a set of candidate formal theories.

Thus

$$\mathcal{E}_A(q_{\mathrm{BH}}, \mathcal{K}_A) \supset \{T_{\mathrm{reaction}}, T_{\mathrm{transport}}, T_{\mathrm{collapse}}\}, \qquad (103)$$

and

$$\mathcal{E}_A(q_{\mathrm{V}}, \mathcal{K}_A) \supset \{T_{\mathrm{contrast}}, T_{\mathrm{flux}}, T_{\mathrm{web}}\}. \qquad (104)$$

Definition 12. Human selection operator

Let

$$\mathcal{S}_H : \mathcal{P}(\mathfrak{T}) \longrightarrow \mathfrak{T} \cup \{\bot\} \qquad (105)$$

select, reject, or redirect machine-generated formalizations.

The symbol

$$\bot \qquad (106)$$

denotes rejection.

Examples include the judgments:

"Not beautiful", (107)
"Too similar to what we have already done", (108)

and

"The physics is disconnected, but the math suggests." (109)

These are not merely stylistic comments. They alter the search trajectory.

Definition 13. Collaborative update

The joint discovery step is

$$\mathcal{D} = \mathcal{S}_H \circ \mathcal{E}_A \circ \mathcal{P}_H. \qquad (110)$$

Iteration gives

$$T_{n+1} = \mathcal{D}(T_n, I_n). \qquad (111)$$

The collaborative process is therefore not

$$H \to A \qquad (112)$$

or

$$A \to H, \qquad (113)$$

but the feedback system

$$\boxed{H \to A \to H \to A \to \cdots .} \qquad (114)$$

Proposition 6. Evaluation is part of discovery

Suppose the machine generates candidate theories

$$T_1, \ldots, T_k. \qquad (115)$$

If human selection changes which candidate is further developed, then the selection operator contributes causally to the final theory.

Proof

Without selection, all candidates remain equally active or the machine must apply an internally chosen criterion. Human rejection changes the subsequent state of the collaborative process. Therefore evaluation is not external commentary; it is an operative part of discovery. ■

The discovery cannot be attributed solely to the generator of formulas.

The rejecting and redirecting acts are equally causal.

8. The Metacognitive Loop

The original discovery process concerned a black disk.

The later observation concerned the discovery process itself.

This introduces a second-order state:

$$M_n = \text{model of the current reasoning process at stage } n. \qquad (116)$$

The first-order sequence was

$$I \to T_{\mathrm{BH}} \to T_{\mathrm{V}}. \qquad (117)$$

The second-order observation was

"A black disk generated black holes, then cosmic voids." (118)

This statement treats the first-order trajectory as a new object of study.

Definition 14. Metacognitive lift

Let

$$\mathcal{L}: (I, T_1, \dots, T_n) \longrightarrow M \qquad (119)$$

map a discovery history into a model M of that history.

For the present episode,

$$\mathcal{L}(I, T_{\mathrm{BH}}, T_{\mathrm{V}}) = T_{\mathrm{DA}}, \qquad (120)$$

where

$$T_{\mathrm{DA}} = \text{the theory of diagrammatic abduction.} \qquad (121)$$

The discovery process therefore generated a theory of its own generation:

$$\boxed{\text{mathematics} \longrightarrow \text{metamathematics of discovery.}} \qquad (122)$$

Definition 15. Reflexive discovery

A discovery process is reflexive when its history becomes an input to a subsequent theory:

$$\mathcal{H}_n = (I_0, T_1, \dots, T_n), \qquad (123)$$
$$T_{n+1} = \mathcal{F}(\mathcal{H}_n). \qquad (124)$$

The present paper is such a reflexive product.

The initial image generated physical mathematics.

The sequence of physical interpretations generated cognitive mathematics.

Proposition 7. Reflexive lift increases abstraction

Let T_n describe an object-level domain and let T_{n+1}describe the process by which T_n was produced. Then T_{n+1} quantifies over structures containing T_n as an element or historical state.

Proof

The object theory treats domain entities as its principal objects. The metacognitive theory treats object theories, interpretations, and transformations among them as its objects. Its domain therefore lies at a higher descriptive level. ■

The abstraction hierarchy is

$$\begin{aligned} L_0 &\quad \text{:tuple image,} \\ L_1 &\quad \text{:black-hole model,} \\ L_2 &\quad \text{:void model,} \\ L_3 &\quad \text{:model of the transition } L_0 \to L_1 \to L_2. \end{aligned} \qquad (125)$$

The transition to $L_3$is not automatic.

It required the human observation that the sequence itself was remarkable.

Definition 16. Metacognitive trigger

A metacognitive trigger is an utterance or event that redirects attention from the object of reasoning to the reasoning trajectory.

The trigger in the present case was:

"The physics is disconnected, but the math suggests." (126)

This sentence separates two levels:

physical validity (127)

from

mathematical generativity. (128)

That distinction initiated the present theory.

9. Diagrammatic Abduction

Classical deduction has the form

$$R,\ C \Longrightarrow O, \qquad (129)$$

where R is a rule, C a case, and O an observation.

Induction seeks a general rule from repeated cases:

$$(C_1, O_1), \ldots, (C_n, O_n) \Longrightarrow R. \qquad (130)$$

Abduction proposes a possible rule or world that would make an observation intelligible:

$$O \Longrightarrow \text{possibly } (R, C). \qquad (131)$$

The present process is diagrammatic because the observation is a visual configuration:

$$O = \mathcal{M}(D_t). \qquad (132)$$

Definition 17. Diagrammatic abduction

Diagrammatic abduction is the generation of a candidate formal theory T from a perceived visual morphology m:

$$m \rightsquigarrow T, \qquad (133)$$

where

$$\rightsquigarrow \qquad (134)$$

denotes conjectural generation rather than logical implication.

Thus

$$m \Rightarrow\!\!\!/\, T, \qquad (135)$$

but

$$m \rightsquigarrow T. \qquad (136)$$

The distinction is essential.

The diagram does not prove the theory.

It proposes a space in which a theory may be constructed.

Definition 18. Abductive validity

An abductive proposal $m \rightsquigarrow T$ is productive when:

$$\begin{array}{ll} 1. & T \text{ is internally coherent,} \\ 2. & T \text{ yields nontrivial deductions,} \\ 3. & T \text{ distinguishes analogy from identity,} \\ 4. & T \text{ states its assumptions explicitly,} \\ 5. & T \text{ survives independent mathematical checking.} \end{array} \qquad (137)$$

Productivity does not require that

$$m$$

literally depict the ontology of T.

Proposition 8. Diagrammatic origin does not invalidate formal consequences

Suppose a theory T is suggested by an unreliable analogy but is subsequently defined axiomatically. Any theorem derived validly from those axioms remains mathematically valid relative to the axioms.

Proof

The validity of a derivation depends on the axioms and inference rules, not on the psychological origin of the axioms. An analogy may motivate a theory without serving as evidence for its physical truth. ■

This yields the distinction

$$\boxed{\text{origin of conjecture} \neq \text{justification of theorem.}} \qquad (138)$$

The black disk supplied the origin.

Conservation proofs and finite-state theorems supplied mathematical justification.

Physical comparison supplied correction and scope limitation.

Definition 19. Abductive distance

Let $m$be a morphology and T a candidate theory.

Define an informal abductive distance

$$d_{\mathrm{A}}(m, T) \qquad (139)$$

as the amount of additional structure required to transform the observed morphology into the theory.

A literal interpretation has small apparent distance:

$$d_{\mathrm{A}}(\text{black disk}, \text{black object}) \approx 0. \qquad (140)$$

A corrected physical theory has greater distance:

$$d_{\mathrm{A}}(\text{black disk}, \text{Einstein–matter collapse}) \gg 0. \qquad (141)$$

Large abductive distance is not necessarily a defect.

It may indicate that the image has served as a remote generator rather than as a representation.

Definition 20. Formal return

After constructing T, one may map its consequences back into a new diagram:

$$T \longrightarrow I_T. \qquad (142)$$

The complete cycle is

$$\boxed{I \to \mathcal{M}(I) \dashrightarrow T \to I_T.} \qquad (143)$$

The new image I_T need not resemble the original image I.

It may instead visualize the newly extracted mathematics:

$$\text{flux,compactness,density contrast,critical barriers.} \qquad (144)$$

Thus diagrammatic abduction can be cyclic:

$$\text{image} \rightarrow \text{theory} \rightarrow \text{new image} \rightarrow \text{new theory.} \qquad (145)$$

10. Morphology Before Ontology

The central philosophical claim is

$$\boxed{\text{morphology precedes ontology.}} \qquad (146)$$

The claim does not mean that morphology determines ontology.

It means that structural perception may occur before any stable assignment of what the perceived object is.

Let

$$m = \mu(I) \qquad (147)$$

be the extracted morphology of an image I.

Let

$$\Omega_1(m), \dots, \Omega_k(m) \qquad (148)$$

be candidate ontological interpretations.

Then the chronology of discovery is

$$I \longrightarrow m \longrightarrow \{\Omega_1(m), \dots, \Omega_k(m)\}. \qquad (149)$$

The morphology is held fixed while ontology branches.

Definition 21. Ontological underdetermination

A morphology $m$is ontologically underdetermined when there exist distinct candidate worlds

$$W_1 \neq W_2 \qquad (150)$$

such that both satisfy the observable constraints imposed by m.

Formally,

$$m \in \mathrm{Obs}(W_1) \cap \mathrm{Obs}(W_2), \qquad (151)$$

where $\mathrm{Obs}(W)$ is the set of observations compatible with W.

The moving black disk was underdetermined because

$$m_{\mathrm{disk}} \in \mathrm{Obs}(W_{\mathrm{BH}}) \cap \mathrm{Obs}(W_{\mathrm{V}}) \cap \mathrm{Obs}(W_{\mathrm{mask}}) \cap \cdots . \qquad (152)$$

Definition 22. Morphological equivalence

Two objects $A$and $B$are morphologically equivalent relative to a feature set $\mathcal{F}$ when

$$f(A) = f(B)\text{for all } f \in \mathcal{F}. \qquad (153)$$

Write

$$A \equiv_{\mathcal{F}} B. \qquad (154)$$

A rendered disk and an astrophysical illustration may be equivalent relative to

$$\mathcal{F} = \{\text{darkness,boundedness,circularity}\}, \qquad (155)$$

while remaining ontologically unrelated.

Thus

$$A \equiv_{\mathcal{F}} B \qquad (156)$$

does not imply

$$A = B \qquad (157)$$

or even

$$\mathrm{Type}(A) = \mathrm{Type}(B). \qquad (158)$$

Proposition 9. Morphological equivalence is weaker than theoretical equivalence

If

$$A \equiv_{\mathcal{F}} B, \qquad (159)$$

then the theories associated with $A$and Bneed not share laws, state spaces, or invariants.

Proof

Morphological equivalence preserves only the selected features in $\mathcal{F}$. The governing theories may depend on unrepresented properties outside $\mathcal{F}$, including causality, density, topology, dynamics, or scale. Therefore morphology alone does not establish theoretical equivalence. ■

The relevant logical failure is

$$A \equiv_{\mathcal{F}} B \Longrightarrow\!\!\!/\, T_A \cong T_B. \qquad (160)$$

Nevertheless, morphological equivalence may motivate a search for a shared abstraction:

$$A \equiv_{\mathcal{F}} B \rightsquigarrow \mathrm{Core}(T_A, T_B). \qquad (161)$$

This is precisely what occurred.

Black holes and voids are not theoretically equivalent.

Their comparison exposed the common language of:

region,boundary,flux,interior,accessibility,threshold. (162)

11. Discovery by Contrast

The second interpretation did not merely add another example.

It transformed the first theory by contrast.

Before the void branch, the dark region was interpreted primarily through

inward concentration. (163)

After the void branch, the same regional formalism admitted

outward evacuation. (164)

The pair produced a signed transport variable

$$\Phi_B = \Phi_B^{\text{out}} - \Phi_B^{\text{in}}. \qquad (165)$$

Then

$$\Phi_B < 0 \qquad (166)$$

indicates net accumulation, while

$$\Phi_B > 0 \qquad (167)$$

indicates net evacuation.

The duality appeared only after both branches existed.

Definition 23. Contrastive discovery

Contrastive discovery occurs when comparison of theories T_1 and T_2 reveals a variable, invariant, or distinction that is not explicit in either theory considered alone.

Write

$$T_1 \bowtie T_2 \longrightarrow J, \qquad (168)$$

where J is the contrastively extracted structure.

For the present case,

$$T_{\text{BH}} \bowtie T_{\text{V}} \longrightarrow \Phi_B. \qquad (169)$$

The regional flux variable unifies the two branches without identifying them.

Proposition 10. Oppositional theories may generate a common signed parameter

Suppose $T_1$and T_2describe opposite changes of the same conserved regional quantity $E(B)$. Then there exists a signed increment

$$\Delta E(B) \qquad (170)$$

whose sign distinguishes the theories.

Proof

If T_1 increases $E(B)$and T_2 decreases $E(B)$, then

$$\Delta E(B) > 0$$

in one theory and

$$\Delta E(B) < 0$$

in the other. Defining

$$\Phi_B = -\Delta E(B)$$

gives a common signed transport parameter. ■

Thus the physics remained disconnected, but the mathematics became connected through orientation:

collapse and void formation are not identical processes; they are opposite directions of regional redistribution.	(171)

Definition 24. Dualization by sign reversal

Let a regional update be

$$E_{t+1}(B) = E_t(B) - \Phi_t(B). \qquad (172)$$

Define the sign-reversal operator

$$\mathfrak{D}: \Phi_t(B) \mapsto -\Phi_t(B). \qquad (173)$$

Then

$$\mathfrak{D}$$

exchanges net evacuation and net accumulation.

This does not exchange black holes and voids as physical objects.

It exchanges one mathematical component of their toy models.

Hence

physical duality is false;flux duality is valid. (174)

12. Productive Disconnection

Ordinary analogy seeks similarity.

The present episode gained structure from dissimilarity.

Let

$$d_{\text{phys}}(T_{\text{BH}}, T_{\text{V}}) \qquad (175)$$

measure physical disconnection, and let

$$s_{\text{math}}(T_{\text{BH}}, T_{\text{V}}) \qquad (176)$$

measure shared formal structure.

The episode exhibited

$$d_{\text{phys}} \gg 0 \qquad (177)$$

and simultaneously

$$s_{\text{math}} > 0. \qquad (178)$$

Definition 25. Productive disconnection

Two theories T_1, T_2 are productively disconnected when

$$d_{\text{domain}}(T_1, T_2)$$

is large, but

$$\text{Core}(T_1, T_2) \neq \emptyset. \qquad (179)$$

The nonempty shared core permits mathematical transfer without ontological identification.

For the present pair,

$$\mathrm{Core}(T_{\mathrm{BH}}, T_{\mathrm{V}}) = \{\text{regional conservation,flux,thresholds,boundary classification}\}. \quad (180)$$

Definition 26. Illicit transfer

A transfer from T_1 to T_2 is illicit when a branch-specific property is treated as though it belonged to the shared core.

For example,

$$\text{causal inaccessibility} \quad (181)$$

belongs to the black-hole branch but not to the void branch.

Likewise,

$$\text{negative density contrast} \quad (182)$$

belongs to the void branch but not necessarily to the black-hole branch.

Therefore the inferences

$$\text{dark region} \Longrightarrow \text{causal horizon} \quad (183)$$

and

$$\text{dark region} \Longrightarrow \text{underdensity} \quad (184)$$

are both illicit without additional premises.

Proposition 11. Shared abstraction permits restricted transfer

Let

$$J \in \mathrm{Core}(T_1, T_2). \quad (185)$$

Then results depending only on J may be transferred between the theories, while results depending on branch-specific structure may not.

Proof

A proof using only J remains meaningful wherever J is instantiated. A proof using additional structures not present in the second theory lacks the required premises there. ■

The valid transfer was the conservation identity

$$\Delta E(B) = -\Phi_B. \qquad (186)$$

The invalid transfer would have been

$$\text{void boundary} = \text{event horizon}. \qquad (187)$$

The metacognitive insight can therefore be stated precisely:

disconnected domains may share a formal skeleton, but only after branch-specific meanings are removed.	(188)

13. The Role of Aesthetic Salience

The black disk attracted attention before it attracted explanation.

This suggests that aesthetic or perceptual salience may act as a search heuristic.

Let

$$\sigma(I, x) \qquad (189)$$

measure the salience of region x in image I.

A region becomes cognitively selected when

$$\sigma(I, x) \geq \sigma_*. \qquad (190)$$

The selected region is then passed to interpretation:

$$x \longmapsto \mu(I\,|_x) \longmapsto \Omega(\mu(I\,|_x)). \qquad (191)$$

Definition 27. Salience gate

The salience gate is the map

$$\mathcal{S}_\sigma : I \longmapsto \{x \in X : \sigma(I, x) \geq \sigma_*\}. \qquad (192)$$

Only regions passing this gate enter deeper abductive processing.

In the present case, the black disk had high salience because of:

contrast, motion, boundedness, violation of background regularity. (193)

Proposition 12. Salience is neither truth nor randomness

A salient feature need not correspond to a physically meaningful entity, but it is not equivalent to arbitrary selection if it systematically identifies structural discontinuities.

Proof

Salience may be generated by contrast or novelty rather than truth. However, contrast and novelty identify deviations from an established visual pattern. Such deviations can provide nonrandom candidates for interpretation even when their first interpretations are false. ■

Thus

$$\text{salience} \not\Longrightarrow \text{truth}, \tag{194}$$

but

$$\text{salience} \longrightarrow \text{candidate for formal attention.} \tag{195}$$

The black disk was important not because it represented a black hole or a void, but because it interrupted a structured field.

Its generative property was

$$\boxed{\text{structured anomaly.}} \tag{196}$$

Definition 28. Structured anomaly

A structured anomaly is a localized feature A satisfying

$$A \notin \mathcal{R}(I), \tag{197}$$

where $\mathcal{R}(I)$is the dominant regularity class of image I, while itself possessing internal organization.

The black disk differed from the background but retained:

$$\text{shape,boundary,trajectory.} \tag{198}$$

It was therefore not visual noise.

It was a compact violation of an existing pattern.

Such violations are natural seeds for abduction.

14. Search Without Premature Ontology

The episode suggests a revised discovery protocol.

Instead of immediately asking

"What is this?" (199)

one first asks

"What structural relations are visible?" (200)

Let the observed feature be represented initially by a neutral descriptor

$$z = (\text{bounded}, \text{moving}, \text{dark}, \text{field-displacing}). \quad (201)$$

Candidate ontologies are then introduced only after the neutral morphology has been recorded.

Definition 29. Ontology-delayed reasoning

Ontology-delayed reasoning separates the discovery process into:

1. observation,
2. morphological encoding,
3. candidate-world generation, (202)
4. formalization,
5. validation.

The order is

$$I \to m \to \{W_i\} \to \{T_i\} \to \{\text{Verdict}(T_i)\}. \quad (203)$$

This prevents the first semantic association from becoming an unexamined premise.

Definition 30. Interpretation ledger

For each candidate world W_i, maintain a ledger

$$L_i = (m, W_i, A_i, D_i, V_i), \quad (204)$$

where:

$$A_i = \text{admitted assumptions,} \quad (205)$$
$$D_i = \text{derived consequences,} \quad (206)$$

and

$$V_i = \text{validation status.} \quad (207)$$

The black-hole ledger would distinguish:

$$\text{visual darkness} \quad (208)$$

from the admitted compactness and causal assumptions.

The void ledger would distinguish:

$$\text{visual emptiness} \quad (209)$$

from the admitted density-contrast and transport assumptions.

Proposition 13. Ontology delay reduces equivocation

If morphology and ontology are represented separately, then a property of the image cannot be silently substituted for a property of the candidate physical world.

Proof

The representation contains distinct variables for observed features and interpreted entities. Any passage from one to the other requires an explicit mapping. Therefore equivocation between “dark in the image” and “physically hidden” or “physically underdense” becomes detectable. ■

The recommended rule is

$$\boxed{\text{name the shape before naming the world.}} \quad (210)$$

15. Failure Modes of Diagrammatic Abduction

Diagrammatic abduction is productive only when analogy remains distinguishable from evidence.

Several failure modes recur.

Definition 31. Literalization error

A literalization error occurs when a visual resemblance

$$A \equiv_{\mathcal{F}} B \qquad (211)$$

is promoted to ontological identity:

$$A = B. \qquad (212)$$

For the present case,

$$\text{dark disk} \equiv_{\mathcal{F}} \text{black-hole illustration} \qquad (213)$$

relative to

$$\mathcal{F} = \{\text{darkness,boundedness,circularity}\}, \qquad (214)$$

but this does not imply

$$\text{dark disk} = \text{black hole.} \qquad (215)$$

Likewise,

$$\text{dark disk} \equiv_{\mathcal{F}} \text{void map} \qquad (216)$$

does not imply

$$\text{dark disk} = \text{cosmic void.} \qquad (217)$$

Definition 32. Feature leakage

Feature leakage occurs when a property belonging to the source image is silently transferred to the target theory.

Examples include:

$$\text{black pixel value} \longmapsto \text{absence of physical energy,} \tag{218}$$

or

$$\text{circular rendered boundary} \longmapsto \text{spherical physical symmetry.} \tag{219}$$

The transferred feature may be merely a rendering convention.

Definition 33. Ontological overreach

Ontological overreach occurs when a formal model is presented as a description of reality without sufficient empirical or theoretical support.

The finite laws

$$M + \bar{M} \longrightarrow R \tag{220}$$

and

$$\mathcal{C} \geq \mathcal{C}_{\mathrm{h}} \longrightarrow H \tag{221}$$

may define a coherent toy system.

They do not establish that the moving disk was generated by annihilation or gravitational collapse.

Similarly,

$$\delta \leq \delta_{\mathrm{v}} \longrightarrow V \tag{222}$$

does not establish that the image represents an astrophysical void.

Definition 34. Formal camouflage

Formal camouflage occurs when mathematical notation gives an analogy an appearance of physical authority without adding genuine constraints.

A symbol such as

$$\mathcal{C} \tag{223}$$

becomes physically meaningful only after:

$$\text{variables,units,domain,measurement rule,validation} \qquad (224)$$

have been specified.

Notation alone does not convert metaphor into theory.

Proposition 14. Formalization does not guarantee empirical relevance

A system may be mathematically coherent and physically irrelevant.

Proof

Internal consistency concerns relations among admitted assumptions. Empirical relevance concerns correspondence between those assumptions and observed reality. The former does not entail the latter. ■

Hence three verdicts must remain separate:

$$\boxed{\text{mathematically valid, physically plausible, empirically supported.}} \qquad (225)$$

Definition 35. Abductive hallucination

An abductive hallucination occurs when a candidate world is generated from a visual cue but its assumptions, derivations, and validation status are not tracked separately.

The danger is greatest when

$$\text{fluency} \gg \text{verification.} \qquad (226)$$

In human–AI collaboration, linguistic coherence may conceal unsupported transfer.

Definition 36. Correction checkpoint

A correction checkpoint is a stage at which the current theory is tested against independent domain knowledge.

Let

$$\mathcal{V}_{\text{math}} \qquad (227)$$

denote mathematical verification and

$$\mathcal{V}_{\text{domain}} \qquad (228)$$

denote domain validation.

A candidate theory should be assigned the pair

$$\text{Status}(T) = (\mathcal{V}_{\text{math}}(T), \mathcal{V}_{\text{domain}}(T)). \qquad (229)$$

Possible outcomes include

$$(\text{pass, fail}), \qquad (230)$$

meaning that the formal system is coherent but the proposed physical interpretation is unsupported.

This is the appropriate status for many productive toy models.

16. A Verifiable Discovery Protocol

The episode suggests an auditable protocol for image-driven mathematical discovery.

Let

$$I_0 \qquad (231)$$

be the initiating image.

Stage 1. Neutral observation

Record only directly observable morphology:

$$m_0 = \mu(I_0). \qquad (232)$$

Avoid physical names.

For the present image,

$$m_0 = (\text{bounded dark disk}, \text{continuous motion}, \text{structured background}, \text{local occlusion}). \qquad (233)$$

Stage 2. Candidate generation

Construct a finite candidate set

$$\Omega(m_0) = \{W_1, \dots, W_k\}. \qquad (234)$$

Examples include:

$$W_1 = \text{absorbing dynamics}, \qquad (235)$$
$$W_2 = \text{causal trapping}, \qquad (236)$$
$$W_3 = \text{underdensity}, \qquad (237)$$
$$W_4 = \text{masked information}. \qquad (238)$$

Stage 3. Assumption isolation

For every candidate W_i, construct an explicit assumption set

$$A_i = \{a_{i1}, \dots, a_{in_i}\}. \qquad (239)$$

No assumption may be attributed to the image unless it is directly observable.

Thus

"the disk is dark" (240)

may be observational, while

"the disk traps causal signals" (241)

is admitted.

Stage 4. Formal derivation

Construct a theory

$$T_i = \text{Closure}(A_i) \qquad (242)$$

under specified inference rules.

Record the distinction

$$A_i \cap D_i = \emptyset, \qquad (243)$$

where D_i is the set of derived results.

Stage 5. Cross-interpretation comparison

Compute

$$\mathrm{Core}(T_i, T_j) \qquad (244)$$

for selected pairs.

This identifies structures that survive ontological divergence.

For the present theories,

$$\mathrm{Core}(T_{\mathrm{BH}}, T_{\mathrm{V}}) \supset \{E, \Phi, B, \partial B, \text{threshold}\}. \qquad (245)$$

Stage 6. External correction

Apply domain-specific constraints:

$$T_i' = \Re(T_i, \mathcal{C}_i). \qquad (246)$$

Record rejected claims separately:

$$R_i = T_i \setminus T_i'. \qquad (247)$$

Stage 7. Return to the diagram

Construct a new visualization

$$I_i' = \mathrm{Render}(T_i'). \qquad (248)$$

The new image should display derived variables rather than merely repeat the initiating appearance.

Stage 8. Metacognitive audit

Record the discovery history

$$\mathcal{H} = (I_0, m_0, \{W_i\}, \{A_i\}, \{T_i\}, \{T_i'\}). \qquad (249)$$

The history itself may become the subject of subsequent analysis.

Definition 37. Provenance-complete discovery

A discovery is provenance-complete when every formal claim is labeled as one of:

$$\boxed{\text{observed,assumed,derived,analogical,externally validated,rejected.}} \qquad (250)$$

Proposition 15. Provenance separation limits analogical contamination

If every claim has a unique provenance label, then an analogical claim cannot be used as a derived theorem without an explicit relabeling step.

Proof

The inference system distinguishes claim classes. Moving a statement from the analogical class into the derived class requires either proof from admitted assumptions or external validation. Silent promotion is therefore detectable. ■

The protocol can be summarized by

$$\boxed{\text{observe neutrally;branch freely;formalize explicitly;correct aggressively.}} \qquad (251)$$

17. Machine-Readable Metacognition

The discovery process may be represented as a finite directed graph.

Let

$$\mathfrak{G} = (\mathfrak{N}, \mathfrak{E}) \qquad (252)$$

be a provenance graph.

Its node classes are:

$$\mathfrak{N} = \mathfrak{N}_{\mathrm{O}} \sqcup \mathfrak{N}_{\mathrm{M}} \sqcup \mathfrak{N}_{\mathrm{A}} \sqcup \mathfrak{N}_{\mathrm{D}} \sqcup \mathfrak{N}_{\mathrm{V}} \sqcup \mathfrak{N}_{\mathrm{R}}, \qquad (253)$$

representing:

$$\begin{aligned} &\mathfrak{N}_{\mathrm{O}} && \text{:observations,} \\ &\mathfrak{N}_{\mathrm{M}} && \text{:morphologies,} \\ &\mathfrak{N}_{\mathrm{A}} && \text{:assumptions and analogies,} \\ &\mathfrak{N}_{\mathrm{D}} && \text{:derivations,} \\ &\mathfrak{N}_{\mathrm{V}} && \text{:validation results,} \\ &\mathfrak{N}_{\mathrm{R}} && \text{:rejections and revisions.} \end{aligned} \qquad (254)$$

Edges are typed:

$$e \in \{\text{extracts,suggests,assumes,derives,contradicts,revises,generalizes}\}. \qquad (255)$$

The initiating path is

$$I_0 \overset{\text{extracts}}{\to} m_0 \overset{\text{suggests}}{\to} W_{\mathrm{BH}} \overset{\text{formalizes}}{\to} T_{\mathrm{BH}}. \qquad (256)$$

The second path is

$$I_0 \overset{\text{extracts}}{\to} m_0 \overset{\text{suggests}}{\to} W_{\mathrm{V}} \overset{\text{formalizes}}{\to} T_{\mathrm{V}}. \qquad (257)$$

The metacognitive path is

$$(T_{\mathrm{BH}}, T_{\mathrm{V}}) \overset{\text{generalizes}}{\to} T_{\mathrm{DA}}. \qquad (258)$$

Definition 38. Discovery depth

The depth of a node $n \in \mathfrak{N}$is the length of the longest directed path from an observation node to n:

$$d(n) = \max\{\ell : I \rightsquigarrow n \text{ by a path of length } \ell\}. \qquad (259)$$

The image has depth

$$d(I_0) = 0. \qquad (260)$$

The physical toy models have greater depth, and the metacognitive theory has greater depth still.

Definition 39. Reflexive edge

An edge is reflexive when its source encodes a discovery process and its target is a theory of that process.

The edge

$$\mathcal{H} \overset{\text{generalizes}}{\to} T_{\mathrm{DA}} \qquad (261)$$

is reflexive.

Definition 40. Revision centrality

For a node n, define revision centrality by

$$c_{\mathrm{R}}(n) = \#\{\text{revision paths passing through } n\}. \qquad (262)$$

A rejected analogy may have high revision centrality even if it is absent from the final theory.

Thus false claims may be historically central without being logically admitted.

Proposition 16. Logical importance and generative importance differ

A rejected statement may have zero logical status in the final theory but positive generative importance in its provenance graph.

Proof

Logical status depends on whether the statement is retained as an assumption or theorem. Generative importance depends on whether the statement caused later distinctions, corrections, or branches. These criteria are independent. ■

The naive identity

$$\text{black disk} = \text{black hole} \qquad (263)$$

was logically rejected but generatively central.

The same is true of

$$\text{black disk} = \text{cosmic void}. \qquad (264)$$

This motivates two records:

$$\boxed{\text{proof graph} \neq \text{discovery graph.}} \qquad (265)$$

The proof graph contains only admitted premises and valid derivations.

The discovery graph also contains failed conjectures, aesthetic judgments, and redirections.

18. Implications for Mathematical AI

A mathematical AI designed only to preserve valid deductions would miss much of the present discovery.

The productive steps included:

visual attention,remote analogy,failed identity,human rejection,cross-domain comparison. (266)

Therefore a discovery-oriented system requires more than a theorem prover.

Definition 41. Dual-ledger mathematical AI

A dual-ledger system maintains:

$$L_{\text{proof}} \quad (267)$$

and

$$L_{\text{discovery}}. \quad (268)$$

The proof ledger stores:

definitions, assumptions, theorems, verification. (269)

The discovery ledger stores:

images,analogies, rejected paths, aesthetic judgments, metacognitive triggers. (270)

The ledgers are linked but not merged.

Definition 42. Branch-preserving search

A search process is branch-preserving when it retains multiple candidate interpretations before selecting one:

$$m \longrightarrow \{T_1, \dots, T_k\} \quad (271)$$

rather than

$$m \longrightarrow T_1. \quad (272)$$

Premature collapse to one interpretation reduces the probability of contrastive discovery.

Definition 43. Correction-seeking objective

Let

$$Q(T) \qquad (273)$$

measure internal coherence and

$$M(T) \qquad (274)$$

measure mismatch with external constraints.

A purely confirmatory system maximizes

$$Q(T). \qquad (275)$$

A correction-seeking system also searches for informative mismatch:

$$\text{maximize}[Q(T) + \lambda G_{\Re}(T)], \qquad (276)$$

where $G_{\Re}$is the structural gain produced by correction.

The goal is not error for its own sake.

It is the discovery of errors whose repair yields new distinctions.

Definition 44. Human veto channel

A human veto channel is an explicit operation

$$\mathcal{V}_H(T) \in \{\text{retain,revise,reject}\}. \qquad (277)$$

In the present work, statements such as

$$\text{“not beautiful”} \qquad (278)$$

and

$$\text{“too similar”} \qquad (279)$$

acted as search constraints unavailable from mathematical verification alone.

A mathematical AI that ignores such judgments may optimize the wrong objective.

Proposition 17. Formal correctness is not a complete search objective

There may exist formally correct theories T_1, T_2 such that a human investigator assigns different scientific or aesthetic value to them.

Proof

Formal correctness concerns validity relative to assumptions. Scientific value may depend on novelty, explanatory power, relevance, or visual salience. These are not determined solely by correctness. ■

The collaborative objective is therefore multi-criteria:

$$\mathcal{J}(T) = w_1 Q_{\text{proof}} + w_2 Q_{\text{novelty}} + w_3 Q_{\text{explanation}} + w_4 Q_{\text{aesthetic}} + w_5 Q_{\text{audit}}. \qquad (280)$$

The weights are not fixed universally.

They are partly supplied by the human investigator.

19. Principal Thesis

The initiating black disk did not encode black-hole physics.

It did not encode cosmic-void physics.

It encoded a morphology capable of supporting several candidate interpretations.

The first interpretation generated a finite theory of:

$$\text{annihilation,radiation,compactness,trapping.} \qquad (281)$$

The second generated a finite theory of:

$$\text{underdensity,outward transport,walls,filaments.} \qquad (282)$$

The physical theories remained distinct.

Their comparison exposed the shared formal skeleton:

$$\boxed{E_{t+1}(B) - E_t(B) = -\Phi_t(B).} \qquad (283)$$

The sign of regional flux then separated the two mathematical branches:

$$\Phi_t(B) < 0 \Longrightarrow \text{accumulation,} \qquad (284)$$
$$\Phi_t(B) > 0 \Longrightarrow \text{evacuation.} \qquad (285)$$

The decisive discovery was therefore not

$$\text{black hole} \leftrightarrow \text{void.} \qquad (286)$$

It was

$$\boxed{\text{one conserved regional law admits opposite directed regimes.}} \qquad (287)$$

The image supplied the occasion.

Analogy supplied candidate worlds.

Correction supplied distinctions.

Comparison supplied the invariant.

Metacognition supplied the present theory.

20. Conclusion

The discovery cycle may be written

$$\boxed{I \to \mathcal{M}(I) \rightsquigarrow \{W_i\} \to \{T_i\} \to \{\Re(T_i)\} \to \text{Core} \to T_{\text{meta}}.} \qquad (288)$$

The stages are:

$$\begin{aligned} I &\quad \text{:image,} \\ \mathcal{M}(I) &\quad \text{:neutral morphology,} \\ W_i &\quad \text{:candidate world,} \\ T_i &\quad \text{:formal theory,} \\ \Re(T_i) &\quad \text{:corrected theory,} \\ \text{Core} &\quad \text{:shared invariant,} \\ T_{\text{meta}} &\quad \text{:theory of the discovery process.} \end{aligned} \qquad (289)$$

The key distinctions are

$$\boxed{\text{resemblance} \neq \text{identity,}} \qquad (290)$$

$$\boxed{\text{conjectural origin} \neq \text{mathematical justification},} \tag{291}$$

and

$$\boxed{\text{physical disconnection} \neq \text{absence of shared formal structure.}} \tag{292}$$

The black disk was first treated as an object.

It then became a generator of candidate worlds.

Finally, the history of those worlds became a new mathematical object.

Thus the complete reflexive sequence is

$$\boxed{\text{diagram} \rightarrow \text{theory} \rightarrow \text{contrast} \rightarrow \text{metatheory.}} \tag{293}$$

The central thesis is:

$$\boxed{\begin{array}{c}\text{Morphology precedes ontology; correction reveals structure;} \\ \text{metacognition converts discovery into mathematics.}\end{array}} \tag{294}$$

21. References

1. Peirce, C. S. *Collected Papers of Charles Sanders Peirce*. Volumes 1–8. Harvard University Press, Cambridge, Massachusetts, 1931–1958.
2. Peirce, C. S. “Pragmatism as the Logic of Abduction.” In *The Essential Peirce: Selected Philosophical Writings*, Volume 2. Indiana University Press, Bloomington, 1998.
3. Hanson, N. R. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge University Press, Cambridge, 1958.
4. Harman, G. H. “The Inference to the Best Explanation.” *The Philosophical Review* **74**, 88–95, 1965.
5. Magnani, L. *Abduction, Reason, and Science: Processes of Discovery and Explanation*. Kluwer Academic/Plenum Publishers, New York, 2001.
6. Magnani, L. *The Abductive Structure of Scientific Creativity: An Essay on the Ecology of Cognition*. Springer, Cham, 2017.

7. Nersessian, N. J. *Creating Scientific Concepts*. MIT Press, Cambridge, Massachusetts, 2008.

8. Gooding, D. C. *Experiment and the Making of Meaning: Human Agency in Scientific Observation and Experiment*. Kluwer Academic Publishers, Dordrecht, 1990.

9. Hadamard, J. *The Psychology of Invention in the Mathematical Field*. Princeton University Press, Princeton, 1945.

10. Pólya, G. *How to Solve It: A New Aspect of Mathematical Method*. Princeton University Press, Princeton, 1945.

11. Pólya, G. *Mathematics and Plausible Reasoning*. Volumes I–II. Princeton University Press, Princeton, 1954.

12. Giaquinto, M. *Visual Thinking in Mathematics: An Epistemological Study*. Oxford University Press, Oxford, 2007.

13. Shin, S.-J. *The Logical Status of Diagrams*. Cambridge University Press, Cambridge, 1994.

14. Barwise, J., and Etchemendy, J. *Hyperproof*. CSLI Publications, Stanford, 1994.

15. Netz, R. *The Shaping of Deduction in Greek Mathematics: A Study in Cognitive History*. Cambridge University Press, Cambridge, 1999.

16. Arnheim, R. *Visual Thinking*. University of California Press, Berkeley, 1969.

17. Kosslyn, S. M. *Image and Brain: The Resolution of the Imagery Debate*. MIT Press, Cambridge, Massachusetts, 1994.

18. Arcavi, A. "The Role of Visual Representations in the Learning of Mathematics." *Educational Studies in Mathematics* **52**, 215–241, 2003.

19. Lakoff, G., and Núñez, R. E. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. Basic Books, New York, 2000.

20. Gentner, D. "Structure-Mapping: A Theoretical Framework for Analogy." *Cognitive Science* **7**, 155–170, 1983.

21. Gentner, D., and Markman, A. B. "Structure Mapping in Analogy and Similarity." *American Psychologist* **52**, 45–56, 1997.

22. Holyoak, K. J., and Thagard, P. *Mental Leaps: Analogy in Creative Thought*. MIT Press, Cambridge, Massachusetts, 1995.

23. Schön, D. A. “Generative Metaphor: A Perspective on Problem-Setting in Social Policy.” In A. Ortony, ed., *Metaphor and Thought*. Cambridge University Press, Cambridge, 1979.

24. Finke, R. A., Ward, T. B., and Smith, S. M. *Creative Cognition: Theory, Research, and Applications*. MIT Press, Cambridge, Massachusetts, 1992.

25. Dunbar, K. “How Scientists Think: On-Line Creativity and Conceptual Change in Science.” In T. B. Ward, S. M. Smith, and J. Vaid, eds., *Creative Thought: An Investigation of Conceptual Structures and Processes*. American Psychological Association, Washington, D.C., 1997.

26. Thagard, P. *Computational Philosophy of Science*. MIT Press, Cambridge, Massachusetts, 1988.

27. Simon, H. A. *The Sciences of the Artificial*. Third edition. MIT Press, Cambridge, Massachusetts, 1996.

28. Boden, M. A. *The Creative Mind: Myths and Mechanisms*. Second edition. Routledge, London, 2004.

29. Hofstadter, D. R., and Sander, E. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books, New York, 2013.

30. Chalmers, D. J., French, R. M., and Hofstadter, D. R. “High-Level Perception, Representation, and Analogy: A Critique of Artificial Intelligence Methodology.” *Journal of Experimental and Theoretical Artificial Intelligence* **4**, 185–211, 1992.

31. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., and Horvitz, E. “Guidelines for Human–AI Interaction.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, 2019.

32. Kamar, E. “Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence.” In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016.

33. Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. “Hybrid Intelligence.” *Business & Information Systems Engineering* **61**, 637–643, 2019.

34. Sutton, R. S. “The Bitter Lesson.” 2019. The essay contrasts manually inserted human knowledge with computational methods capable of discovering structure through scalable search.

35. Wigner, E. P. “The Unreasonable Effectiveness of Mathematics in the Natural Sciences.” *Communications on Pure and Applied Mathematics* **13**, 1–14, 1960.

36. Penrose, R. *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford University Press, Oxford, 1989.

37. Peirce, C. S. “The Logic of Drawing History from Ancient Documents, Especially from Testimonies.” In *The Essential Peirce: Selected Philosophical Writings*, Volume 2. Indiana University Press, Bloomington, 1998.

38. Youvan, D. C. *From Annihilation to Horizons: A Finite Algebraic Toy Model for Reactive Matter, Radiation, and Gravitational Collapse*. Unpublished manuscript, 2026.

39. Youvan, D. C. *From Horizons to Voids: A Two-Barrier Finite Algebra of Cosmic Structure*. Unpublished manuscript, 2026.

40. Youvan, D. C. *Tuple Imagery*. Visual and computational project, 2011–2026.